(51) International Patent Classification [7] :		(11) International Publication Number:	**WO 00/62242**
G06K 9/20	A1	(43) International Publication Date:	19 October 2000 (19.10.00)

(21) International Application Number: PCT/BG00/00010

(22) International Filing Date: 5 April 2000 (05.04.00)

(30) Priority Data:
103323   9 April 1999 (09.04.99)   BG
103505   18 June 1999 (18.06.99)   BG

(71)(72) Applicant and Inventor: POPOV, Ivaylo Nicolaev [BG/BG]; J.k."Mladost 2", Bl.211, Vh.2, Et.3, Ap.27, 1799 Sofia (BG).

(81) Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
With amended claims and statement.

(54) Title: METHOD FOR HUMAN–MACHINE INTERFACE BY DOCUMENTS

(57) Abstract

It is represented a method for human–machine interface by documents, which are standard described with a local descriptor that is graphically written in the same document and a global descriptor of document and its relations with other documents disposed in a global accessible data base. The document is recognized using a sample described with the descriptors and juxtaposed to the document using a special graphic image written on each page of the document. A standardized document certification gives the opportunity for secured transport of information through electronic or graphic way using documents. This allows full automation of the document processing.

Method for human-machine interface by documents

Field of techniques

The present invention relates to data input in a machine by recognition of graphically written data within a document by means of a special standardized structure of the document graphic image. It relates to automatic determination of the center of the image coordinate system, the vector of translation and the angle of image rotation by analysis of a special image written in a reserved area on every page of the document. It relates to data transfer by documents through a globally described document structure. It is relevant to automatic document processing. It is relevant to a certification of documents.

Background of the invention

The following patents are most near to background of this invention:

U.S. PATENT DOCUMENTS

5,305,396 4/1994 Betts, et al.

5,307,423 4/1994 Gupta, et al.

5,625,721 4/1997 Lopresti, et al.

5,745,610 4/1998 Johnson

5,815,160 9/1998 Kikuchi, et al.

The Patent No 5,305,396 discloses how to extract data from a graphical original of a document in accordance with a template. It applies semantic interpretation of the document fields.

The methods using segment symbols to write information on any postal piece and bank checks are known. For example, the U.S. Patent No 5,307,423 offers to use well-known seven-segments outlines for numbers and pattern "British flag" for characters. The separate strings are localized through bar codes and other

markers. The use of this method is related to standardization of the respective string positions on postal piece and bank checks.

There are known some methods, which provide incorporation of special machine-readable information together with the original document information into the same document for recognition verification (U.S. Patent No 5,625,721). The mentioned patent gives an example for a non-human readable record of digital information by a graphic image. It is not proposed any method for creation and recognition of such image. It is proposed a digital graphic record, which contains data for the document structure, the structure information kind and the special area certificate. The graphically written digital information is not related to any standardization of a document recognition structure. The certificate should be ciphered to determine the document authenticity. It is not proposed any method for standard certification through a certificate center that keeps the deciphering information. It is not given any certificate for arbitrary graphic image. The certificate is not used for recognition of a graphic image.

There are some methods for document marking with goal to link the documents to a database (U.S. Patent 5,745,610).

This text contains information previously applied at the Patent Office of Republic of Bulgaria - application No 103323 dated April 09 1999.

Internet - the global information network is in background of the invention. In the beginning of the nineties the network technologies were intensive developed owing to the idea to write a standardized document descriptor together with the document with goal to can the document be interpreted on different platforms and to can link to the document a procedure described knowledge. Some authors are naming this local document descriptor scenario by analogy with the theatre and the same term in the artificial intelligence because the local descriptor really gives a scenario for document processing.

The transport technologies of information for last years were advanced and at this moment the interactive connection is background of the invention. That allows to be made the next step in document processing namely the standardization of document information structure and procedure described knowledge for each document and for relations among documents by disposition of the standardizing information on a globally accessible network site.

The new in the offered method is that is given a general method for recognition of graphically human marked structures forming characters, digits, choice fields etc., printed text, hand made text, graphic images, graphically written digital information, etc. by a standardized structure of the document descriptor in physical and logical level and writing a part of it by a graphic digital record in the same document and disposing of the other part on a globally accessible data base.

This gives the opportunity for creation of different forms of a document and its recognition by standardization of the document information structure. This standardization can be used for documents in electronic form too, which allows totally standardization of the documents and relations among them and make it possible the full automation of document processing.

Summary of the invention

A document means whichever unit of information written on whichever kind of data medium, using whichever method in way that can be read using whichever apparatus and can be interpreted as a graphic image. Within a document are defined some areas in logical level and special graphic images assigned to them. The set of these areas furthers recognition of documents by a machine no matter what kind of information a document contains - handwritten or entirely made by machine. Conditionally these areas can be divided into areas with control

information and data areas. The data areas can be filled in by hand or machine and can be recognized. The areas with control information are divided into areas for determining physical parameters and areas determining logical parameters. The data areas are divided into ordinary and special data areas. The special data areas are used to correct hand filled in data. The data areas are defined by a structure and being recognized according to their structural marked status and graphical defined, that can be recognized juxtaposing a recursive defined structure of the sample and a dynamic relative made structure for recognition of real image. Every one of the areas is defined by a digital graphical record into the same document. The digital graphical record is made by structurally defined data. To establish congruence between the graphical image of a document and the sample used for recognition of the same document it is written a special image on every page of the document. To determine the authenticity of a document it is used a certificate that is made according to the method of Lopresti that must be ciphered. The document structure is standardized by a global descriptor containing standardizing information disposed in a global accessible database. The certification of the documents is standardized by creation of certificate centers that contain independent or customer information for creation and ciphering-deciphering of document certificates. The communication with the certificate centers is performed using a secured communications channel.

Detailed description

1. There are sensitive areas in the document.

The sensitive areas represent areas of the image, described by their closed contour. The sensitive areas are of different kinds in relation with the information that these areas contain and with the way to process it. The sensitive areas are segments

of the graphical image. The sensitive areas can be included in a graphical figure printed in the document.

A value in the two-dimensional scanning plane is defined for every pixel of the sensitive area. This value indicates the pixel color or the level of gray.

For each kind of sensitive area it is defined a procedure described method for sensitive area processing.

1.1 A processing method for marked sensitive areas.

The values of pixels are summed up and it is found a normalized value by a function in relation to the number of pixels. The value interval is fixed and in this way is juxtaposed to a defined data. The fixed value interval determines the value of data. In this way can be written different kinds of data as binary data, hexadecimal data and so on. There can be different values of marking, if it is made using different level of gray, different color and different number pixels of the same color and background pixels of other color. Simply the marking is binary. There is a defined level over that an area is considered as marked. The binary marking is used to write a sign by hand fill in a segment structure, to mark a choice and to record binary information.

1.2 A processing method for sensitive areas containing fixed images as printed text, a photo, etc.

When the graphical image is of an exact defined set, the recognition must be made by juxtaposing the examined image to images in the set and fixing the maximum similar image in the set. This juxtaposing can be made directly or using the method further disclosed for comparing image certificates. A sample is looked for which certificate has the nearest value to the certificate value of the examined image.

1.3 A processing method for very deformed images as hand made text, a signature, looking for similar segments in a random image.

6

The recognition must be implemented like recognition of recursive defined sensitive areas to minimize the influence of deformations. It is defined for every recognizable image an unique recursive defined structure of sensitive areas that contains a descriptor of surrounding contour, a function for making up a certificate of the area or a data list controlling an universal certificate function, a function for relative dynamic positioning of the area in the examined image.

A recognition process examines the recursive nested sensitive areas for a deformed image. The position recognition of sensitive areas in the examined image is ordered in way that every next area can be defined in relation to the information about previous examined areas.

The defined sensitive area of a structure for recognition of handmade text or a deformed graphic image is always considered as first.

For all sensitive areas the image is normalized, it is made up a certificate using the appropriate way for the examined area and these certificates are juxtaposed to sample certificates with goal to evaluate deviation of the image from its sample. It is chosen a sample that has minimal deviation.

It is possible to create some methods for automatic definition of a recursive recognition structure: a function for positioning the areas in the image, a function for creating a certificate of the area and a normalizing function for images in sensitive areas in the examined image and the sample.

The normalizing function can not only center but also can evaluate the scale rate and angle of the image rotation in a sensitive area through juxtaposing the size and position of fixed sensitive areas in the examined image to the sample.

The function for area positioning in the image is used for dynamic relative positioning of an area in relation with the information for previous recognized areas. This function represents a special method for determining and multiple

checking of all possible areas, which are determined by an appropriate method, to obtain absolute maximum of approximation between the sample area certificate and the examined area certificate.

2. A structural sensitive area represents a spatial defined structure of sensitive areas. The structural sensitive areas are defined by a list of descriptors of sensitive areas. The sensitive areas descriptor can be stored in the memory of the recognition computer or can be graphically written in the same document.

For every kind of structural sensitive areas it is determined a procedure described method for processing it.

2.1 The value of structural sensitive area for marking is evaluated from the values of sensitive areas regarding their order in the list. The sensitive areas values are interpreted as digits of a number that is in number system with radix N, where N is the number of possible different values for every sensitive area there is a relation between the number and a determined data type, which is stored in a data base, positioned in a recognizing computer or graphically written in the document. If it is examined a machine record of digital information, the resulting number will represent the same data.

The evaluation of a number according to previous discussed method and using it as a key in a database is a real recognition process for a sign in accordance with the marked status of structural area. It is possible to figure more from one number juxtaposed to one sign in a database. In this way, all possible markings in the structure, which determine the same sign, are defined.

2.2 The value of structural sensitive area for fixed images represents a chosen image from a set that is best juxtaposed to the image in the area.

In this way the minimum of evaluated deviation between the examined image and a sample is found. When is used the method

for recognition by juxtaposing certificates of the examined image and a sample, the deviation value is the absolute value of difference between these two certificates. The certificates of different sensitive areas are made up using a special method for processing each area, defining weight of the pixels in this area. A sample with minimum deviation from the examined image is found.

2.3 The value of a structural sensitive area for deformed image is a sample image that is nearest to the deformed image.

In accordance with this method, by the function for positioning of a sensitive area in the examined image, it is found possible disposition of each one sensitive area for fixed image in the examined image. The images in all areas are normalized and their certificates are calculated by respective procedures to each one area. It is calculated the summary deviation of the respective pair of certificates. It is found a sample with minimum summary deviation.

3. The set of structural areas is a spatial defined set of structural areas.

The set of structural areas contains some structures for recognition of a digital graphical record and some structures for data recognition. The structures for data recognition can be of each one of the mentioned types above.

The set of structural areas has a set of values. The digital graphical record writes it a set of digital values in a defined format in order to describe the structure for recognition of the set and disposition of the structural areas in the document, as well as a pointer to a certificate or a ciphered certificate of the set using Lopresti's method for certificate creation.

The structural data areas form a group of structural areas. Each group of structural areas has a value that is made up from the values of separate structures. Into the set of structural areas can be defined a set of groups of structural areas, which means the set of structural areas can obtain a set of the same

type data as a result. The data item is simple determined by the number of that group of structural areas which result is.

The recognition of structural areas set begins by recognizing a digital graphic descriptor of the set. In accordance with data contained in the record for logical and physical structure of structural areas set it is dynamic determined the recognition process. As a result it is obtained a list of data, which type is determined by the document logical structure descriptor and a pointer to next set of structural areas.

In the logical level the document graphical image can be recognized by juxtaposing it to an image model that is described by a list of structural areas set. A pointer to the first item of this list for each page of the document is in standard positioned digital record that means a label of the page. At least the label of first page contains a standardized digital record determining digital record characteristics of a document.

4. The recognition of a special image written on each page of the document by juxtaposing the document to an image model is the first step of the recognition process.

The special image written on each page of a document consists of a set of concentric circles like a "shooting-mark", each one of them is written with different thickness. The center of circles set is the center of image model coordinate system, too. These circles can be interpreted as a generalization of the bar code. Outside of the circles set a straight line segment is written on X-axis direction of the image model, that is named X-marker or on Y-axis direction of the image model that is named Y-marker.

The set of circles is examined through intersection by a straight line. The straight-line equation is used as a functional relation between coordinates of straight-line points and in this way it is determined an examination trajectory for this sign. There is an intersection point with the sign, when a point that is found using the straight-line equation is dark.

The distance between end intersection points is watched. Different trajectories for moving the scanning straight line are implemented. It is possible to rotate the scanning straight line relative to the center of scanning plane coordinate system; it is possible to move it in X-axis direction being parallel to Y-axis. In this way it is found a scanning straight line that has the longest distance between end intersection points with sign. Through the point that determines the middle of this distance it is drew a fictive straight line at a determined angle with previously found line.

The sets of intersection points of these straight lines with the sign are compared. There must be a coincidence of all these points regarding to their order and value in the perfect case. There are a maximal number of discrepancies under that the sign is considered as recognized in the general case. An additional check up is made for coincidence of circle thickness to previously defined set of combinations, so each one of these combinations determines different coordinates of initial digital record and different size and form of sensitive areas building the structure of initial digital record. The choice of relation between circle thickness and size of their radiuses is standardized for all documents, regarding standard sizes of digital record areas as well as their standard positions and forms, regarding possible resolution of used printers and scanners. If the sign was not recognized, it is possible to determine another trajectory for scanning straight line and another angle between the scanning straight line and a correcting straight line. That can be made with goal to avoid any physical destruction of the sign. If the sign was recognized using previously discussed method, the image model center is exactly the point that determines the middle of distance between two end intersection points of the straight line and the sign. That is so because of definition cause the image model center of coordinate system is equal to the center of scanned circles. The

evaluation of image scale is made by comparing the expecting as it is in the model and real number of intersection points of the scanning straight line and sign. The physical data for possible configurations of the sign are a part of standardized information. Through found image center is fictive drew a scanning straight line that is parallel to X-axis of scanning plane. If the image is not rotated, the scanning straight line must contain the X-marker. If the image is rotated in process of printing or scanning, the X-marker must be looked for in the image moving a scanning straight line parallel to Y-axis of scanning plane in the area of X-marker. If an intersection point was found, a new scanning straight line is fictive drew through the image center and the found intersection point. It is checked up if this straight line contains the X-marker (it is possible that the intersection point is a "noise"). Else the first scanning straight line is taken away from the center. For real case it is not necessary the scanning straight line to contain completely the X-marker. It is determined a number of X-marker points that must be contained in the scanning straight line (X-marker may be partially destroyed). In this way it is evaluated the angle of image rotation around the image model center. The same is valid for Y-marker, but the scanning straight line is parallel to X-axis of scanning plane. The image translation vector was found when the image model center was found in the scanning plane. A coincidence between each one point from the image model and point/group of points from the scanning plane is determined by the evaluated characteristics. The group of points is examined as fictive sensitive area and its value is examined instead of the image model pixel.

Following functional relations are defined:

$Xs = Ftx(Xm)$

$Ys = Fty(Ym)$;

$VAL(Xs,Ys) = Fm(Xij,Yij)$; i =1.. n; j:=1..m;

Where

Xm, Ym are coordinates of an image model point.

Xs, Ys are coordinates of a non-scaled point of scanning plane.

Ftx, Fty are transforming functions for the respective coordinates.

Fm is a function that evaluates the value VAL of point (Xs,Ys), using the appropriate point set.

These functions are juxtaposing a point from the image to a point / a group of points of the scanning plane through rotation, translation, and scaling. The coordinates that are graphically written into a document with their absolute value and are not relative are in image model coordinate system.

5. A coordinate description of the area.

Every structural area is defined in a rectangular area of the document and its coordinates are the coordinates of upper left corner of the rectangular area. The sensitive area coordinates are relative to the coordinates of structural areas, where they are positioned. The structural area coordinates are relative to coordinates of the structural areas set where the areas are positioned. Those relative coordinates mean coordinates expressed in following way: [a (X2-X1), b (Y2-Y1), where a and b are coefficients, greater than or equal to zero and minor than or equal to one, and X1, Y1, X2, Y2 are coordinates of surrounding rectangular area of the structure. The using of these coordinates in description of nested structures make it possible a relative change of nested structure coordinates in relation to coordinates and size of the structure, containing them.

6. A digital graphical record.

In a special structural areas set it is written a number in numeric system with radix N and the addition to N-1 of the sum of the number digits by module N.

The spatial disposed digital graphic record provides further information about the correct recognition by check-up of digital record correctness using a control digit. If the distortions are wrong corrected or are nonlinear (for example the paper is crumpled) it is possible some graphical descriptors cannot be read. It means that the field data cannot be read correctly, too.

The set of structural areas for digital graphical record can be used also for establishment of correct position during the recognition process and it can have suitable for this purpose structure.

For detection of nonlinear distortions it can be disposed also some special digital records in some defined places of the document, independently from the digital field descriptors. The information that is obtained from them can be used for further correction. It is possible to process a minimal correction of coordinates of inaccurately recognized structures, to be compensated the distortions.

7. An area for error correction during a hand marking.

The area for error correction is a part of above described structure. This area is described such as a group of structural areas sets that must be corrected. A structuring identifier from a global nomenclature defines the area for error correction. By means of the structuring identifier it substitutes the structure that is marked appropriately in the area for error marking with a structure from the correction area marked through the same way in the area for error correction. The area for error marking can represent an area for binary graphic record by a hand marking. This area is used such as a label of corresponding structural areas.

8. Standardizing the logical document structure and methods for their processing by a global descriptor.

Within a digital graphical record it is written information for logical and physical structure as well as control

information. The information for logical structure may contain identifying and characterizing data for a document field. The structuring identifiers from a global nomenclature, those determine the recognition process represent logical information too. In particular, such structuring identifiers are of following types - document type, field type, information structure type or nonstructural image type. An advantage of structuring identifiers is that they make it possible in every moment to be included in a database a new structuring identifier together with its sense (procedure described knowledge). All those allow the new recognition method to be step by step added.

The logical document structure is also defined through a global accessible database of documents. This base may be disposed on a global accessible site through a network. It may be also a local copy of the global accessible base.

The global base must contain:

- A nomenclature of structuring identifiers that are interpreted as pointers to procedure described knowledge or to data;

- A documents descriptor according to document types nomenclature, containing a list of descriptors of document fields and document fields function relations;

- A descriptor of relations between the fields of one document and the fields of another document, and of the functional transformations, which must be applied according with these relations. With a view of this descriptor there are different structuring identifiers by which is determined different sense of the given document during its processing;

- A nomenclature of standard information that contains a list of standard fonts, standard structures, standard sizes and another.

The document is associated with its descriptor through a structuring identifier from the document types nomenclature that is written in initial digital record. The descriptor can be

local or global. The local document descriptor has the same structure as the structure of global descriptor, but it is applied for processing of those documents that are used only by the user itself. The global descriptor is standardized all over the territory where it is used. Documents that are created according with the global descriptor can be used for data transfer between all users conforming to this descriptor. That provides an opportunity to create systems for completely automatic documents processing.

The global descriptor does not specify the disposition of document fields within the document or size of these fields, but it specifies the area of initial digital record and the special area with information for accordance between an image and a recognition model.

The description of document physical structure and other characteristic control information is written in the graphical original of document through a digital graphical record. It is possible the physical structure descriptor to contain also structuring identifiers for recognition process control.

It is not necessary the digital graphical record of logical structure information to be different from the digital graphical record of physical structure information. Just, in the digital graphical record it must be written in addition information for a format of this record. In other words, the digital graphical record contains information about its logical and physical structure.

The document standardization includes also the standardization of the resulting list that is formed during recognition process. The resulting list contains also a certificate.

9. A standardization of certificate issuance of documents.

9.1 A method for certification of the documents.

A certificate of the document is made up according to the method, disclosed in U.S. Patent 5,625,721 Lopresti et al., but

this certificate must be ciphered. It is certified only the information in sensitive areas. For text areas it is certified the recognized text independently from its recording manner into the document.

For fixed graphical image a certificate is made according to a defined weight for each pixel of the image. For example, more important are pixels in the border between two colors in the image. Together with the certificate it is written a value that means the acceptable deviation of previously made certificate from secondary made certificate.

For deformed graphic images it is found the determining image contour that is defined as combination of all border pixels, surrounded with defined number of neighboring pixels with different colors.

The contour is represented as black and white graphic image, where the black pixels forming the contour are weightier in certification process, and white pixels have zero weight. It is possible to define some sets of contour points having one and the same weight. In this way prepared the information is named contour map of the sample.

It is created a certificate of the sample, evaluating the sum of pixels weights in sample contour map.

It is created a contour map of sample similarities in way that for each pixel with zero weight is determined a weight evaluated from distances to the nearest pixels with weight and their weight.

It is created a contour map of the examined image, finding the pixels weight in contour map of sample similarities. It is created a certificate of the examined image summing weights of pixels in the examined image contour map. This is the certificate of the examined image contour map.

The certificate of the deformed image contour map means as the image certificate.

An admissible deviation value may be written together with the certificate. When is looked for an image in a known set, it is looked for a sample that has the minimum certificate deviation.

9.2 A method for standardization of the certificate.

It is known, that the certification cannot be free processed. The man that is issued a certificate will know the deciphering information and can verify the documents.

This means that for document verification it must be always looked for its publisher.

It is possible to make an automatic certificate center that is globally accessible. It is possible to use in that center two methods for certification. The first method is to save in a database deciphering information for certificates issued from registered customers of the center. The second method is to be issued from the center an independent document certificate. When there is an independent document certification, the center receives a document image or its resulting list through a secured communications channel. In the center a document is certified using the same method, as the center was the publisher of document. The center may use different ciphering information for each document and store it in a database by the time that is previously fixed. The certificate center receives a resulting list from document recognition or a document graphic image. The certificate center examines the certificate that is composed from identifying not ciphered part and a ciphered part that is composed from an identifying part and a document certificate. Who customer was issued the document is found using the identifying part. The necessary deciphering information is extracted from the deciphering information base. It is compared if non-ciphered identifying information is equal to the ciphered identifying information. It is made a new document certificate on the resulting list or on the document graphic image. The new document certificate and the deciphered certificate are compared

and if the deviation is admissible, the certificate center confirms the document authenticity. When a document is independently certified, its identifying part contains a registration number of certification, the date and time of certification. Using this information it is found deciphering information for the document certificate. The certification is as customer certification. The independent certification may be used for electronic notaries certification.

Brief description of the drawings

Fig.1 illustrates some various sensitive areas.

Fig.2 illustrates a segment structure for digital input by marking.

Fig.3 illustrates a graphically formed structure for digital information input by marking.

Fig.4 illustrates a segment structure for input of symbols by marking.

Fig.5 illustrates graphically formed structure for input of symbols by marking.

Fig.6 illustrates various kind of marking the same number.

Fig.7 illustrates a special structure for digital graphical record.

Fig.8 illustrates a set of structural areas for digital graphical record.

Fig. 9 illustrates a set of structural areas for input of digital information by marking and special areas for error marking 91.

Fig. 10 illustrates the special image that is used to juxtapose the sample to the recognizable image.

Fig.11 presents in detail the image of Fig.10.

Fig.12 illustrates the moving of an image in relation to the scanning plane.

Fig.13 illustrates the geometry meaning of the method for center determination, rotation angle determination and translation vector determination of the image.

Fig.14 is a view of a generalized scheme of the method for standardization of the document structure and document processing.

Fig.15 presents the method in a diagram.

Detailed description of the preferred embodiment

1. In a document can be defined by an arbitrary closed contour various sensitive areas. For example, in the document are defined four cornered sensitive areas and they are parallelograms. In fig.1 it is illustrated six example areas. The sensitive areas can be included into a graphical figure, disposed in the document (fig.3 or fig.5). The sensitive areas 11 and 12 are spatially disposed to be formed the seven-segment model of number into the graphical figure 3. Each sensitive area is a segment of the document image.

1.1 For purpose of the example, the sensitive area for marking is examined through the sheaf parallel to one of tetragon's sides' lines. The equation of line is processed as a functional relation between coordinates of pixels of examining trajectory. Each pixel has a defined value in the two-dimensional scanning plane. In accordance with above described method the sum of pixel values is normalized in relation to the number of pixels, for example it is evaluated an average sum value. This value determines the logical value of sensitive area by finding the value interval that contains it. For binary values it is defined a constant that determines the boundary between dark and light and divides the set of colors into two

non-intersection intervals (equal for example). So the areas for binary marking have two states - marked or non-marked (1,0).

1.2 For example, the sensitive area for fixed image A and sample B are processed under the formula (NOT A AND B) OR (NOT B AND A), the values of resulting image pixels are summed, and it is found a sample for which the obtained result is most near to 0 (zero). The colors of examined image must be prior made in accordance to the colors of the sample.

1.3 For the example, it is examined a sensitive area for deformed image under recognition of handwritten capital "A".

The image is divided into four sensitive sub areas. The recognition begins by positioning of first sensitive area in the image. The area that contains the horizontal line of the handwritten sign "A" is determined as the first sensitive sub area. The sign must be scanned by a horizontal line until a straight line that intersects the sign in points which number is greater than the previously defined number is found. For each line as that above it is determined a line segment between the end points obtained by intersection of the line with the sign in its surrounding area, the image in the area is normalized by determining scales of both axes through a comparison between the found sensitive area and the model sensitive area. The image is centered. It is made a certificate of the image by the method for deformed graphic image. It is chosen a sensitive area with nearest certificate to the sample area certificate. It is examined the second sensitive area. The part of image over the first sensitive area must be scanned from left to right by a vertical line until it is found a straight line which intersects the sign in points which number is greater than the previously defined number. It is scanned also on direction from right to left by the same way. It is scanned from top to down by a horizontal line under the same condition. All three found straight lines together with the outline of first sensitive area determine a sensitive area. The image of this area is scaled,

centered and transformed to be compensated the difference between the real sign incline and the sample sign incline. The difference between two inclines is determined by mutually disposition of the sensitive areas or by the cut and try method to obtain the best congruence. The cut and try method can be used to be more accurately determined the sensitive area through a step by step moving of the determining straight lines. The third and the fourth areas are located by the way that the image which is under the first area is scanned from left to right, from right to left, from down to top. It is found a straight line that potentially separates the two areas. This line can be in conformity with the determined at present deformations of sign. It is started a scanning by lines that are parallel to the above discussed separating line in opposite directions. The scanning goes on to the intersection between the scanning line and the sign in points which number is greater than the previously defined number. The third and the fourth areas are processed like the second. The obtained image sub areas certificates deviations are summed and the result represents the total deviation between recognizable image and the sign.

This is only an example for recognition of a deformed image of handwritten capital letter "A". It gives an idea about procedural and structural description of deformed image recognition process. That does not restrict the recognition process. Likewise it is possible to recognize a signature, and more over it is able to examine the features of an unspecified graphics by determining of typical parts of it under the method.

Voice signals for example can be present as graphics and examined under the method.

2. The structural areas represent a spatial defined structure of non-intersecting sensitive areas like these in the fig.2, fig.4 and fig.7. In the fig.7 are presented sensitive areas 10 disposed into a matrix 3x3. The structural areas are defined by

a list of sensitive areas descriptors. The sensitive areas descriptors are of type (X1,Y1,X2,Y2,X3,Y3), where

X1,Y1 - upper left corner

X2,Y2 - bottom right corner

X3,Y3 - upper right corner

The value of a structural sensitive area is determined by the values of all sensitive areas according with their order in the list.

2.1 The value of structural sensitive area forms a number. It is associated a defined type data with this number in a database. In the database it is defined all possible structure marking for one sign. The numbers "9" and "6" are showed by two various markings in the fig.6.

2.2 The value of structural area for a fixed image is the recognized fixed image.

2.3 The value of structural area for deformed image is the sample, which certificate is at minimum deviated from the deformed image certificate.

3. A set of structural areas represents spatial defined set of structural areas.

For example, the structural areas set 8 consists of some line disposed structural areas 7 which are used for a digital record. With each one structural areas set it is associated a special set of structural areas consisting of a spatial disposed matrix 3x3 of sensitive areas for write a digital information with a control digit (fig.8). This set is always standard disposed. For example, the initial structural area of the digital record 8 possesses the coordinates of the structural areas set 9 that is marked by it. Fig.9 shows a disposition of the set 9 that consists of one associated set 8, three structural areas disposed within graphical figures 3 and error correction structures 91.

Into the structural areas set it is able to dispose each kind of data structure. The structural areas set descriptor define that.

The descriptor 155 contains the following fields, for example:

- A structure of the physical features descriptor of a digital record. It may be a byte that contains a bit determining the direction of recognition (from top to bottom or from left to right) and a number of the structures that determine the physical features. The direction that is pointed by the bit is the direction of structure recognition for a digital record. The structure recognition order is determined by its structure descriptor according to the method.

The physical features are following:

- Number of structural areas within a group, which areas are disposed in a direction pointed by the special bit from the first byte;

- Number of groups that are disposed in direction opposite to the direction pointed by the special bit;

- Number of structural areas in the same record; - the radix of the number system in which is the digital record (This field is optional and it is binding only in the initial digital record);

- Step X - gives the distance between two structures in the direction of X - axis.

- Step Y -gives the distance between two structures in the direction of Y - axis;

- A number of a certificate for verifying the field or the group of fields that belongs to.

It follows the information part of the digital record:

- Descriptor kind - a number that is structural identifier disclosing about what is the descriptor;

- Kind of structural area or a number of document field - kind of structural area is a number that defines the kind of

structural area that can be a structural area for alphabetic, numbers, choice, graphic image, printed text, handwritten text and another.

- Information for verification of field input values as a set of admissible values or a type of input data. (This field is optional and it is used only when the information in the global document descriptor differs from the information about respective field);

- Kind of recognition structure;

- Number of structural areas within a group - it determines the number of structural areas within the group that are disposed in the direction that is defined in the field "Direction";

- Direction - it defines the direction of structural areas groups' recognition process (left to right or top to bottom);

- Number of the same type groups - it defines the number of the same type groups, for example, for document fields which have many values;

- Step X - it is the distance between two neighboring structural areas in the X-axis direction;

- Step Y - it is the distance between two neighboring structural areas in the Y-axis direction;

- Xmax, Ymax - those define the size of the window of structural areas set imaging, and wherefrom the size of structural areas;

- Page, X, Y - those are the coordinates of next structural areas set.

This information is for purpose of the example. It supposes that the other part of information is into the global descriptor.

At the beginning of each document it is defined an initial structural areas set for a binary information record. The binary record provides some physical characteristics of the digital

record as the radix of number system, kind of a structure, size and etc. After initial binary record recognition it is recognized the initial identification record. The identification record contains, for example, the following fields:

- A document type - this is a structuring identifier from a global nomenclature, that links the document with its global descriptor;

- A document identifier - this is the number of the respective copy of this document kind;

- A document author identifier - it may be, for example, a registration number for companies, an identification number for physical persons or other;

- A document page - it defines the number of the marked page;

- Page, X, Y - These are the coordinates of next structural areas set.

In every document page is positioned a digital record considered as a label, containing:

- A document page - it determines the number of a marked page;

- Page, X, Y -coordinates of next structural areas set.

Every structural areas set can contain sensitive structural areas for marking. Using them it can be formed digits by the seven segment model (fig.2,3), characters and symbols by the special segment model (fig.4,5), as well as indicate a choice by marking.

Every structural areas set can contain rectangular areas with graphical information. In this way, using consecutive rectangular areas containing graphic image it can be defined a recognition process for handmade or printed text. For printed text the descriptor digitally written in a document contains information about size, font and other font characteristics - bold, italic and so on. For printed text recognition it is enough to extract information from an appropriate database, containing the standardized fonts. The received graphic

information is compared with sign images form in an appropriate font table.

To this moment the handmade text meant text, which handmade signs are positioned in separate, graphically written in the document frames. The recognition of handmade text is processed sign by sign.

When is recognized real handmade text, each word in the text separated by spaces must be processed using the method for sign recognition considering that word as first sensitive area and the next structural level contains sensitive areas for handmade signs. The sensitive areas for handmade signs are established by a dynamic cut and try method. The handmade signs are recognized by maximum similarity. The received word is dictionary tested and if it is not correct it is chosen the nearest word from the dictionary, satisfactory conformable to the recognized information. For this goal the recognized characters can be changed with next characters a formed list of similarities for each determined sensitive area for handmade sign.

Recognizing handmade text, the sample set must contain different image kinds for a sign.

Recognizing a signature or fingerprints, the image is examined just as if the whole image is a handmade sign.

4. On each document page it is printed a special image (fig.10, where is shown a special sign 100 and element 101, outside of it), serving to juxtapose a sample with a real scanned image. For this example, the coordinate system center of sample 130 is chosen to be in upper left corner of the image. Into the document graphic image it is printed a set of concentric circles (element 100) with center, coinciding with the sample coordinate system center 130. For example, the maximum size of sign 100 is fixed and standardized. The most outside circle determines the maximum size of the sign and serves to evaluate the scale. The scale is evaluated using the relation between the diameter of a standardized maximum size in

number of pixels and the received number of pixels for the diameter of the real scanned image. This relation may be standardized and the sign may be used for determination of the nearest standard value. In this way it is received exact congruence, independent of the discrete structure of the circles that can cause errors. For same reason, the digital information record using the circles is made with standardized relations between thickness of the circles. Let us assume that the most inside circle has a fixed diameter such as the most outside. The most inside circle is filled with black color, so its radius 111 is equal to its thickness. The thickness of the most inside circle serves as a standard. Every circle between the most inside and the most outside circles can have a number of standard relations with the standard most inside circle. The radiuses 112,113,114 of the circles are determined by a step that is composed from a standard distance (element 115) for circles separation and the thickness of the circles (the thickness are shown as elements 116,117,118). Each circle excepting the most inside determines a digit of a number in numeric system with radix equal to the number of relations between thickness of the circle and the most inside standard circle thickness. A number is formed in this way. The described method may be examined as generalization of bar code and can be independently used, as a way for digital graphic record, everywhere is used bar code. This method is independent from a position in the space and from scale of the scanned image compared with the sample. This allows to be used it for sorting of marked according to this method objects, using a digital camera and image processing for control of sorting process.

The formed number using the described method serves to determine the standard of the initial binary record in the document. This standard includes the size in pixels of squares, the record structure and it standard position in the document.

In fig.12 is shown an example position of the document 120 in the scanning plane 121. In fig. 13 is shown the position of scanning coordinate system 133, the coordinate system of the sample 134, the vector of translation 131, angle of rotation 132, coordinate system center 130, that is center of the sign 100, too. For example it is used X-marker 101 disposed in X-axis.

5. The coordinate description of the areas.

The coordinates of structural area sets are given with their exact values in a number of pixels. The coordinates of structural areas and sensitive areas are for example:

An example descriptor of the rectangular sensitive area is:

[0.3X,0.2Y,0.5X,0.4Y,0.5X,0.2Y]

An example descriptor of the nonrectangular sensitive area is:

[0.3X,0.2Y,0.5X,0.4Y,0.8X,0.2Y]

6. The digital graphical record.

For a digital record can be used rectangular sensitive areas 11 which are a part of space matrix, for example 3X3 (fig.7). This case is examined only as an example but is not a limit. The sensitive areas can be different - described with arbitrary contour and arranged in arbitrary space structure.

For example, the information about the digital record is graphically written by a binary record. All rest records are made in hexadecimal numeric system, supposing that the used hardware can distinguish minimum 256 gray nuances. This enormous coding excess furthers exact delimiting of different color intervals corresponding to the digits.

A structural identifier determines the information structure of the digital record. It is different for recognition of printed text, handmade text or hand marked structural areas, representing a sign.

7.The area for error correction in case of hand marking.

A special structural identifier indicates an area to be an area for error correction. If an error is made during hand marking of a segment structure for numbers or letters input, then it must be marked the specially positioned under the segment marking structure 9 a row of parallelograms 91, which is interpreted as segment input of a binary number. This binary number is a label of the wrong marked structure and of the structure for correction that is marked by the same way in the correction area. Thereby the wrong structure is replaced during recognition process with the structure from the relevant area for correction. It may be defined a rule that it does not allow to mark all parallelograms from the label at marking of a correcting structure. Thereby if all parallelograms of the label are together marked, that means the whole labeled structure for correction must be ignored, in other words this structure is wrong too.

8. The standardization of logical document structure and methods for document processing by a global descriptor.

It is created an Internet site where the global descriptor 140 of documents (fig. 14) is. This site is under the State control and the data that is established on this site is obligatory for all subjects that have some relations with the State. The information of the global descriptor is interpreted as a national standard.

It is possible that individual subjects extract directly information from this Internet site or it is possible that a part of this information is stored on a local carrier of one subject. In the last case the subject is responsible for the actuality of its copy.

8.1. The way of presentation of a resulting list and the resulting list structure are standardized. For example consecutive disposed information structures of type:

8.1.1. The information for record structure:

- Length of the information structure.

- Type of the information structure.

8.1.1.1 The identifying record:

- Type of the document - a structural identifier from a global nomenclature, used for a link between the document and its global descriptor. For example, one type is "Invoice". For the identifier it is not important of what kind is it- textual or numerical. For example it is textual;

- A characterizing identifier - a structural identifier from a global nomenclature that is used to characterize the document. For example, "Invoice for sale", "Invoice for supply", etc.;

- An identifier of document author - a number from national or international register;

- The number of document according to number list of the author;

- The number of document according to number list of the recipient;

- Date and time of document issue;

- Date and time of document receiving;

- Storage time;

- Date and time of last processing;

- A status field of the document, for example "Archived", "Finished", "Frozen", "In process", etc.

8.1.1.2 The information record:

- A field number - a field number from the global document descriptor;

- number of field certificate, that links the field with its certificate;

- Data length;

- The data.

8.1.1.3 The certificate record:

- Length of certificate;

- The certificate.

8.1.1.4 A record of each one structure which type is defined according to 8.1.1.1

8.2 The standardizing information from a global descriptor of the document.

8.2.1 The nomenclature is an indexed data base with following structure:

8.2.1.1 A structure of the nomenclature:

- number of nomenclature, which is the registration number of the nomenclature in the date base;

- A number that is the registration number of the structural identifier in its nomenclature;

- The structural identifier;

- Additional information for the structural identifier. For example, help information.

8.2.1.2 The structure of the nomenclature classifier.

- Nomenclature number that is the registration number of a nomenclature in the database;

- The nomenclature name;

- The nomenclature type, indicating the type of nomenclature's values;

- Length of nomenclature values;

- Other information about the nomenclature as help information.

8.2.2 The structure of a document descriptor:

- A document number from the document nomenclature;

- A document field number;

- The field type that is a structural identifier of the nomenclature for types;

- The field length.

- Number of expected field values.

8.2.3 The structure of a procedure knowledge descriptor:

- A nomenclature number;

- A number of a structural identifier from the nomenclature;

- A number of a structural identifier from the nomenclature for context, using to structures the procedure knowledge. For example "Calculate", "Correct the error", "Certify", etc.;

- A pointer to a procedure.

8.2.4 The structure of oriented relation between a document and another document.

8.2.4.1 The structure of relation:

- A relation number;

- Number of first document from the document nomenclature;

- A number of first characterizing identifier from the context nomenclature;

- Number of second document from the document nomenclature;

- A number of second characterizing identifier from the context nomenclature.

8.2.4.2 The structure of procedure descriptor of the relation between the fields of the two documents:

- A relation number; ·

- A number of a field of the first document;

- A number of structural identifier from the context nomenclature;

- A pointer to a procedure.

8.2.4.3 The procedure has a standard form. For example, it is written in a standard language that can be interpreted by each one computer that is networked. The procedure can call all procedures pointed in which one of described above ways.

This is only an example for a standard structure of information. The standardized information is very more various, but this example illustrates how is it able to describe the structure of documents and knowledge about the relations between them.

9. The standardization of document certification.

Some State licensed Internet sites are created to issue document certificates. By these sites it is made a customer and / or independent (free) certification. The independent

certification is produced in special offices that have a contract with the certificate center. The methods for customer and independent certifications are the same with exception of that in the case of independent certification the office itself is a customer and the certification center creates an official record containing information for deciphering of the document certificate for each document. In case of customer certification the deciphering information is stored into the list of possible values and it is able to certificate a great number of documents through one and the same deciphering information. The certificate center 141 communicates with own customers by Internet connection using security protocol 142. A result list 152 is sent to the certificate center 141 to issue a certificate. It is possible to send the graphical image of the document 153 to the certificate center 141 to make up a certificate. In this case it must be first produced a resulting list 152 of the document in the certificate center 141.

The standardized certificate contains at least for example:

9.1 A non ciphered part - the identification part of the certificate, which contains:

- A certificate center address that is the address of the certificate center answering for the certification;

- A kind of a certificate, disclosing whether the certificate is issued by a registered customer of the certificate center or it is independently issued;

- A number of the document publisher, disclosing a number of a registered customer of the center or an officially generated number for an independent certificate;

- Date and time of the certificate issue.

9.2 The ciphered part of the certificate, which contains:

- A kind of a certificate, disclosing whether the certificate is issued by a registered customer of the certificate center or it is independently issued;

- A number of the document publisher that is a number of a registered customer of the center or an officially generated number for an independent certificate;

- Date and time of the certificate issue;

- A certificate type, disclosing the structure of the certificate and the way for it's processing;

- The maximal admissible distortion, that is maximal admissible difference between the deciphered certificate and the second made certificate;

- The certificate length;

- The certificate.

- A ciphered part with official information of the center.

9.3 The information stored in the center.

9.3.1 The information about registered customers. That is at least:

- A customer number;

- Deciphering information;

- A validity term of the decoding information.

In the database it is possible to be there some such records for a customer.

9.3.2 The information about the independent certificates:

- A number of certification;

- Date and time of certification;

- Deciphering information;

- A validity term of deciphering information.

9.3.3 Official information of the center.

- A list of daily generated ciphering information from the center that is used for ciphering of center information area where it is kept an officially generated certificate of a certificate;

The certified resulting list 150 is verified applying at the center 141 that is issued the document certificate. The address of the center is a part of the certificate. In accordance with the customer identifier or the officially generated number for

an independent certificate it is found the deciphering information about the document certificate. The customer identifier or the officially generated number for an independent certificate is incorporated in the certificate too. The certificate center 141 compares the deciphered document certificate with the certificate that is produced on transmitted graphical image of the document 151 or on a resulting list 150 from recognition process. The two certificates are considered as a list of certificates of rectangular areas from the document, which are calculated using various methods according to the area type. In case of a graphical image area the certificate may be deviated from the defined one no more from a value that is a part of the certificate too. If the comparison of all certificates from the list is successful, it is considered that the document itself is successfully recognized and it is correct. It is reported the areas for which the certificates are non coincident and if it is possible, an operator can correct the recognized information. The correction may be necessary in case of large deviation in recognition process like it is·in handwritten text recognition or printed text recognition.

9.4 The information structure of the certificate.

The document certificate is a list of information structures of following type:

- A number of a certificate, that is an identification number by which it is determined the certified by this certificate fields of the document. That number may be not used if the order number of a structure in the list is substituted for the number of certificate. Let's it used for clearness of the example;

- A certificate type, for example for a fixed image, for a deformed image, for text, for another;

- Length of the certificate;

- The certificate.

The document certificate is treated, for example, as united whole under the execution of procedures that were described above.

Practical applications of the invention

The invention is applicable anywhere where are used documents, allowing their automatic processing. The invention represents a method for description of document recognition structures as well as recognized data processing, their certification and transfer. As a method for description of recognition structure and associated with it knowledge which are necessary for recognition and processing of the documents, this method is able to become a basis of improvement of automatic information processing, because it offers an opportunity for freely including of latest processing methods which are or will be invented, as a part of recognition process.

It should not be seen the documents only as the paper documents. This invention considers the document as a great multitude of ways for writing and reading information that can be interpreted as a graphical image and be treated according to the offered method. This may be a computer graphics, television image, fax. But it is possible to generalize still further if non-graphical information is converted in graphical information and it is treated according to the method. For example, a graphically presented result from some measuring of a dynamic process (like cardiogram) can be recognized by the method. It is possible the graphically presented voice characteristics to be recognized using the method, too.

The opportunity for document certification is especially useful. The certification is the basic problem of electronic transfer of documents. A document which is reliable certificated can be not only a tax, an administrative, an accounting, or an identifying document, but also the rarely understand like

document currency, for example. If it includes a digital record according to the method into its graphical image and this record contains the face value of the bank note and a certificate of graphical bank-note image in addition to the traditional protective items, the bank note can be directly used for electronic trade. This makes the electronic trade very reliable, because it gives the alternative means for payment when the pay - system fails (the payment proceeds traditionally by bank notes, which is absolutely necessary in case of retails. For the bank notes it may be used independent certification according with the method by a certificate center that is the Central bank, for example.

The special image for juxtaposing the sample to the real image is useful too and it can be independently used like an improved analogue of the bar code as well as can be used for centering of arbitrary images or physical positioning of objects.

What is claimed is:

1. A method for human-machine interface by documents that
are recognized by examining of their graphical image;
are juxtaposed to a sample by a special image written on each
page of the document;
are recognized using a structural descriptor of the document;
are transmitted through a communications channel;
are certified, comprising:

a special image written on each document page to juxtapose
the real image to the sample;

a descriptor for each structural area field of the document,
written by digital graphical record;

a structural identifier from a global nomenclature, that is
digitally, graphically written in the document, linking the
document with its global descriptor in logical level;

a ciphered certificate of the document written by digital
graphical record in the document;

any program instructions written in the document by digital
graphical record;

a control of recognition process using a global descriptor
disposed in a global database that can be globally accessible
and using a local graphically written descriptor of recognition
structure with goal to produce:

recognition of hand marked segment structures for input of
characters and digits,

correction of hand filled structures,

recognition of printed text,

recognition of marked choice areas,

recognition of images by special certification,

recognition of handmade text and deformed images,

creation of a resulting list from the recognition process,

transmition to a certificate center of a resulting list
through a secured communications channel together with a digital

ciphered certificate that is printed in the document or of a graphic image for applying a certificate procedure and receiving an answer about authenticity of the recognized document;

data transfer through a standardized logical structure of the document, described in a global descriptor;

automation of recognized data processing;

standardization of documents according to the method to automate theirs processing.

2. The method of claim 1 wherein a special image is written on each page of the document with goal to juxtapose the real image to a sample, comprising the steps of:

writing a set of concentric circles with different thickness and a center that is the center of the sample coordinate system;

printing outside this set of circles a standard direction marker on X-axis or Y-axis.

3. The method of claim 2 wherein a special image is recognized and used for juxtaposing the real image to a sample, comprising the steps of:

tracing out of the circles set with a straight line under a defined angle, until the maximum distance between two end intersection points of the straight line and the image is received;

drawing a correction straight line under a defined angle to the first straight line through the point determining the middle of distance between the end intersection points of the first straight line with the circles set;

comparing of the intersection points of the two straight lines;

finding a correlation between continuous dark segments and / or light segments on the scanning straight line and it interpretation as a number;

using of the received number to find a standard model of a digital graphic record in a database;

tracing out by a straight line that is parallel to perpendicular axis to axis on which is drew the standard marker for direction and using this marker for determination of the rotation angle;

comparing the received number of intersection points with circles set with predefined one for finding the standard sample scale toward the image.

4. The method of claim 1 wherein a document descriptor is created for document recognition, comprising:

a descriptor of sensitive areas;

a descriptor of structural sensitive areas;

a descriptor of structural sensitive areas set.

5. The method of claim 4 wherein a descriptor of sensitive areas is created, comprising:

creating a data list that describes a closed contour surrounding the sensitive area.

6. The method of claim 4, wherein a structural sensitive areas descriptor is created, comprising steps of:

creating a list of sensitive areas descriptors, describing the structural sensitive area.

7. The method of claim 4 for creating a descriptor of structural areas sets, comprising the steps of:

creating a descriptor for each structural areas set;

all descriptors are linked in a linear list according to their physical nearness in the document;

each descriptor contains at least an identifier of the field, type of recognition structure, a certificate number and physical parameters of the structural areas set as area size Xmax, Ymax, number of structural groups, direction of structure recognition, distances through X-axis and Y-axis between two neighboring structures;

in the descriptor can be contained different structural identifiers from a global nomenclature with goal to control the recognition process or processing of respective field.

8. The method of claim 1 wherein a global descriptor is made for facilitation of recognition process and standardization of document structure, comprising:

disposition of a global descriptor of the documents and relationship between them in a global accessible site;

taking into consideration the information in the global database, while a standard document is recognized.

9. The global descriptor structure of claim 8 comprising:

a nomenclature of document structural identifiers, interpreted as pointers to a descriptor of the logical document structure and to procedure described knowledge for document processing;

a nomenclature of structural identifiers, interpreted as pointers to procedure described knowledge or as switches of make up process, or as data;

a composed nomenclature that contains two structural identifiers and indicates an oriented relation between the documents described with these identifiers and possibly with characteristic structural identifiers for the two documents. The composed identifiers from the nomenclature are interpreted as pointers to descriptors of relations between fields of two documents and procedure information for their processing;

a composed nomenclature comprising a structural identifier of the document and a characteristic structural identifier interpreted as pointer to a list of linked documents, i.e. raising documents and engendering documents by the document with the respective characterization and if is fulfilled the procedure described condition of the relation;

a nomenclature of standardized information, interpreted as a pointer to information for standard fonts, standard structures, time limits, sizes, etc.

a nomenclature of structural identifiers, used for structuring the procedure described knowledge;

a nomenclature of identifiers naming the different recognition processes.

10. The method of claim 1 for creating a digital graphical record, comprising:

mono marking with different density or color marking of the structural sensitive area;

receiving a digit as a number of color interval wherein the marking of the sensitive area is;

forming a number by sorting the digits according to order of structural sensitive areas examination;

a initial binary record according to the method characterizing the digital record by the radix of numeric system for the written number, color interval of record and structure characterizations.

11. The method of claim 1 for recognition of marked segment structures, comprising:

interpretation of a structural area as a number according to the method of claim 10;

forming a binary number using it as key in a database containing the respective signs;

in the database more from one number may be juxtaposed to the same sign.

12. The method of claim 1 for correction of hand filled structures comprising:

determining of a correcting structures area for each document page;

associating a structure for binary record by hand marking to data structures and to correcting structure and using it as a label;

using the same marking of the structures labels for finding the correcting structure.

13. The method of claim 1 for recognition of a graphic image by certification comprising steps of:

making up a sample contour map;

weight definition for contour pixels;

forming a contour map certificate by using pixels weight;

forming a similarity map by evaluation of a weight for each pixel of the sample contour map where said pixel has not a weight;

preparing a contour map of the examined image;

weight definition of the contour pixels according to the sample similarity map;

forming a contour map certificate of the examined image by using the pixels weight;

image recognition by comparing its certificate with a previously defined certificate or with image certificates from a defined set;

determination of the minimum deviation between the examined image certificate and an image from the set;

accepting an image as recognized if that image has the maximum near certificate to the examined image certificate;

when a certificate is compared with a previously defined certificate, it is checked if the deviation between two certificates is in previously defined limits.

14. The method of claim 1 for recognition of printed text comprising the steps of:

determining printed text symbols using described sensitive areas in the recognition structure for printed text;

finding through a structural identifier from a standardized nomenclature a set of images that can be juxtaposed to the symbol being recognized according to its characterizations;

normalizing the color of the symbol being recognized and of the sample and thereupon using a method for direct juxtaposing or the method of claim 13.

15. The method of claim 1 for recognition of hand made text with spaces between characters or of deformed graphic image comprising the steps of:

determining handmade signs using sensitive areas from the recognition structure for handmade text or deformed image;

using that the sensitive areas in a document are graphically outlined;

the image from one sensitive area is perceived as a handmade sign or as a deformed image that possibly contains a group of handmade signs;

for each image it is defined a recognition structure of recursive defined sensitive areas that serves as a sample when a recognition structure is dynamic built;

for each sensitive area of the sample it is defined a special method for examining or an ordered pixels list according to their importance, a function for relative position of the area in the examined image, a function for image normalization, a function for certificate making up;

the certificate of each sample sensitive area is compared with its respective certificate of an area from a dynamic built structure, prepared using the special method of the sample area;

determining a sample that has .a certificate with minimum deviation from the certificate of the examined sign according to the recognition structure of the sign.

16. The method of claim 15 for dynamic building of recognition structure by tray and cut method comprising steps of:

examining an initial supposed sensitive area;

coping the image from it;

normalizing the image in the copy of the area;

preparing a certificate of the image copy using the special method for respective sample sensitive area;

comparing the prepared certificate with the sample certificate;

area modifying and executing again the same method on the modified area to receive the absolute maximum of coincidence for the examined image;

17. The method of claim 1 for creating a resulting list from the recognition process comprising:

a linear list of information structures containing a field identifier, a number of field certificate or of a group of fields to which it belongs, size of the field and value of the field;

a part of the resulting list is a list of certificates of fields groups or fields;

the list of certificates is a linear list of information structures containing a certificate number, a certificate type determining a procedure for its making up, the certificate length and the same certificate.

18. The method of claim 1 for certification of a resulting list or of a graphical document image comprising the steps of:

creation of some licensed certificate centers in sites, which are accessible through a global network;

communications through a secured channel between customers and the center;

sending a graphic document image or a resulting list from its recognition to the certificate center for preparation of a document certificate or preparation of a document certificate in a local computer of a registered customer;

ciphering the document certificate in the certificate center or in the local computer using the officially generated information from the center (independent certificate) or using a special database of ciphering information for every registered customer;

printing the formed certificate, using digital graphical record in the same document;

sending the resulting list or the graphical document image to the certificate center when the document is recognized for verification of the recognized information authenticity;

making up certificates for each certified area of the document and their comparison with deciphered certificate that is written in the same document.

19. The ciphered certificate structure of claim 1, comprising:

a non-ciphered identification part;

a ciphered part containing identification information, certificate information and an officially ciphered data certificate of the certificate.

20. The method of claim 1 for data transfer by a standardized document structure comprising the steps of:

creating a resulting list by a recognition process according to the global document descriptor;

sending the resulting list through a communications channel;

in the receiving station the resulting list is checked up by an inquire to the certificate center;

on base of the resulting list it creates a graphical image of the document in accordance with a local descriptor of a graphical document image;

the created graphical image of the document has the same certificate power as the original document, because it can be always examined by an inquiry to the certificate center;

the data transfer can be made up using transmition of a graphical document image through a communications channel and checking up this image by an inquiry to the certificate center.

21. The method of claim 1 for automatic processing of recognized data comprising:

automatic generation of documents from a list that is pointed from the composed nomenclature of structural document identifiers and characteristic structural identifiers according to claim 9;

using the composed nomenclature of structural document identifiers and characteristic structural identifiers according to claim 9 for determination of field links of generated documents with fields of the base document and functional relations between them;

the result of automatic processing is a resulting list for a document that can be transformed in a graphic image according to claim 20.

22. The method of claim 1 for document standardization, comprising:

writing in the document all or some graphical images and using all or some methods with goal to standardize the recognition of the documents in way that a document means each information combination written on whatever carrier using whatever method so that it can be read by whatever apparatus and can be interpreted as an image;

a document in electronic form that is represented by its resulting list is in standard form, too.

23. Some documents of claim 22 are:

text documents;

accounting documents;

administrative and legal documents;

labels of goods;

addresses of parcels;

currency, securities, stocks, postage stamps;

printed publications - books, periodicals;

automatic or hand created forms for handmade input in a machine as game slips, cash-slip checks, voting-papers etc.;

identification documents;

graphic images in cardiograph;

electronic publications - graphic images published through electronic way;

transformed in graphic form non-graphical information processed according to the method.

AMENDED CLAIMS

What is claimed is:

1. A method for human-machine interface by documents that
are recognized by examining of their graphical image;
are juxtaposed to a sample by a special image written on each
page of the document;
are recognized using a structural descriptor of the document;
are transmitted through a communications channel;
are certified, comprising:

a special image written on each document page to juxtapose
the real image to the sample;

a descriptor for each structural area field of the document,
written by digital graphical record in the same document;

a structural identifier from a global nomenclature, that is
digitally, graphically written in the document, linking the
document with its global descriptor in logical level;

a ciphered certificate of the document written by digital
graphical record in the document;

any program instructions written in the document by digital
graphical record;

a control of recognition process using a global descriptor
disposed in a global database that can be globally accessible
and using a local graphically written descriptor of recognition
structure with goal to produce:

recognition of hand marked segment structures for input of
characters and digits,

correction of hand filled structures,

recognition of printed text,

recognition of marked choice areas,

recognition of images by special certification,

recognition of handmade text and deformed images,

creation of a resulting list from the recognition process,

transmitting in first case to a certificate center a
resulting list or a graphic image through a secured
communications channel to receive certified resulting list or
graphic image with digital ciphered certificate or making in the

second case the ciphered certificate in the local computer according to rules that are contracted with a certificate center;

transmitting to a certificate center a certified resulting list or a graphic image with a digital ciphered certificate for applying a certificate procedure and receiving an answer trough a secured channel about the authenticity of the recognized document;

data transfer through a standardized logical structure of the document, described in a global descriptor;

automation of recognized data processing;

standardization of documents according to the method to automate theirs processing.

2. The method wherein a special image is written on each page of the document with goal to juxtapose the real image to a sample, comprising the steps of:

writing a set of concentric circles with different thickness and a center that is the center of the sample coordinate system;

printing outside this set of circles a standard direction marker on X-axis or Y-axis.

3. The method of claim 2 wherein a special image is recognized and used for juxtaposing the real image to a sample, comprising the steps of:

tracing out of the circles set with a straight line under a defined angle, until the maximum distance between two end intersection points of the straight line and the image is received;

fictive drawing a correction straight line under a defined angle to the first straight line through the point determining the middle of distance between the end intersection points of the first straight line with the circles set;

comparing the intersection points of the two straight lines;

finding a correlation between continuous dark segments and / or light segments on the scanning straight line and it interpretation as a number;

using the received number to find a standard model of a digital graphic record in the global descriptor;

tracing out by a straight line that is parallel to perpendicular axis to axis on which is drew the standard marker for direction and using this marker for determination of the rotation angle;

comparing the received number of intersection points with circles set with predefined one for finding the standard sample scale toward the image.

4. The method wherein a local document descriptor is created for document recognition, comprising:

a descriptor of sensitive areas;

a descriptor of structural sensitive areas;

a descriptor of structural sensitive areas set.

5. The method of claim 4 wherein a descriptor of sensitive areas is created, comprising:

creating a data list that describes a closed contour surrounding the sensitive area.

6. The method of claim 4, wherein a structural sensitive areas descriptor is created, comprising steps of:

creating a list of sensitive areas descriptors, describing the structural sensitive area.

7. The method of claim 4 for creating a descriptor of structural areas sets, comprising the steps of:

creating a descriptor for each structural areas set;

all descriptors are linked in a linear list according to their physical nearness in the document;

each descriptor contains at least an identifier of the field, type of recognition structure, a certificate number and physical parameters of the structural areas set as area size Xmax, Ymax, number of structural groups, direction of structure recognition, distances through X-axis and Y-axis between two neighboring structures;

in the descriptor can be contained different structural identifiers from a global nomenclature with goal to control the recognition process or processing of respective field.

8. The method wherein a global descriptor is made for facilitation of recognition process and standardization of document structure, comprising:

disposition of a global descriptor of the documents and relationship between them in a global accessible site;

taking into consideration the information in the global descriptor using a secured channel or using a local probably partial certified copy of the global descriptor, while a standard document is recognized.

9. The global descriptor structure of claim 8 comprising:

a nomenclature of document structural identifiers, interpreted as pointers to a descriptor of the logical document structure and to procedure described knowledge for document processing;

a nomenclature of structural identifiers, interpreted as pointers to procedure described knowledge or as switches of make up process, or as data;

a composed nomenclature that contains two structural identifiers and indicates an oriented relation between the documents described with these identifiers and possibly with characteristic structural identifiers for the two documents. The composed identifiers from the nomenclature are interpreted as pointers to descriptors of relations between fields of two documents and procedure information for their processing;

a composed nomenclature comprising a structural identifier of the document and a characteristic structural identifier interpreted as pointer to a list of linked documents, i.e. raising documents and engendering documents by the document with the respective characterization and if is fulfilled the procedure described condition of the relation;

a nomenclature of standardized information, interpreted as a pointer to information for standard fonts, standard structures, time limits, sizes, etc.

a nomenclature of structural identifiers, used for structuring the procedure described knowledge;

AMENDED SHEET (ARTICLE 19)

a nomenclature of identifiers naming the different recognition processes.

10. The method for creating a digital graphical record, comprising:

mono marking with different density or color marking of the structural sensitive area;

receiving a digit as a number of color interval wherein the marking of the sensitive area is;

forming a number by sorting the digits according to order of structural sensitive areas examination;

using an initial binary record according to the method characterizing the digital record by the radix of numeric system for the written number, color interval of record and structure characterizations.

11. The method for recognition of marked segment structures, comprising:

interpretation of a structural area as a number according to the method of claim 10;

forming a binary number using it as key in a database containing the respective signs;

in the database more from one number may be juxtaposed to the same sign.

12. The method for correction of hand filled structures comprising:

determining of a correcting structures area for each document page;

associating a structure for binary record by hand marking to data structures and to correcting structure and using it as a label;

using the same marking of the structures labels for finding the correcting structure.

13. The method for recognition of a graphic image by certification comprising steps of:

making up a sample contour map;

weight definition for contour pixels;

forming a contour map certificate by using pixels weight;

forming a similarity map by evaluation of a weight for each pixel of the sample contour map where said pixel has not a weight;

preparing a contour map of the examined image;

weight definition of the contour pixels according to the sample similarity map;

forming a contour map certificate of the examined image by using the pixels weight;

image recognition by comparing its certificate with a previously defined certificate or with image certificates from a defined set;

determination of the minimum deviation between the examined image certificate and an image from the set;

accepting an image as recognized if that image has the maximum near certificate to the examined image certificate;

when a certificate is compared with a previously defined certificate, it is checked if the deviation between two certificates is in previously defined limits.

14. The method for recognition of printed text comprising the steps of:

determining printed text symbols using described sensitive areas in the recognition structure for printed text;

finding through a structural identifier from a standardized nomenclature a set of images that can be juxtaposed to the symbol being recognized according to its characterizations;

normalizing the color of the symbol being recognized and of the sample and thereupon using a method for direct juxtaposing or the method of claim 13.

15. The method for recognition of hand made text with spaces between characters or of deformed graphic image comprising the steps of:

determining handmade signs using sensitive areas from the recognition structure for handmade text or deformed image;

using that the sensitive areas in a document are graphically outlined;

the image from one sensitive area is perceived as a handmade sign or as a deformed image that possibly contains a group of handmade signs;

for each image it is defined a recognition structure of recursive defined sensitive areas that serves as a sample when a recognition structure is dynamic built;

for each sensitive area of the sample it is defined a special method for examining or an ordered pixels list according to their importance, a function for relative position of the area in the examined image, a function for image normalization, a function for certificate making up;

the certificate of each sample sensitive area is compared with its respective certificate of an area from a dynamic built structure, prepared using the special method of the sample area;

determining a sample that has a certificate with minimum deviation from the certificate of the examined sign according to the recognition structure of the sign.

16. The method of claim 15 for dynamic building of recognition structure by tray and cut method comprising steps of:

examining an initial supposed sensitive area;

coping the image from it;

normalizing the image in the copy of the area;

preparing a certificate of the image copy using the special method for respective sample sensitive area;

comparing the prepared certificate with the sample certificate;

area modifying and executing again the same method on the modified area to receive the absolute maximum of coincidence for the examined image;

17. The method for information transmittal structure standardizing according to the global descriptor, comprising:

making following structures according to the global descriptor:

a resulting list that means a linear list of information structures containing a field identifier, a number of field

certificate or of a group of fields to which it belongs, size of the field and value of the field;

a part of the resulting list is a list of certificates of fields groups or fields,

where the list of certificates is a linear list of information structures containing a certificate number, a certificate type determining a procedure for its making up, the certificate length and the same certificate.

18. The method for certification of a resulting list or of a graphical document image comprising the steps of:

creation of some licensed certificate centers in sites, which are accessible through a global network;

communications through a secured channel between customers and the center;

sending a graphic document image or a resulting list from its recognition to the certificate center for preparation of a document certificate or preparation of a document certificate in a local computer of a registered customer;

ciphering the document certificate in the certificate center or in the local computer using the officially generated information from the center (independent certificate) or using a special database of ciphering information for every registered customer;

printing the ciphered certificate, using digital graphical record in the same document;

sending the certified resulting list or the graphical document image to the certificate center when the document is recognized for verification of the recognized information authenticity;

making up certificates in the certificate center for each certified area of the document and their comparison with deciphered document certificate;

transmitting an answer about document authenticity from the certificate center to the customer.

19. The ciphered certificate structure, comprising:

a non-ciphered identification part;

a ciphered part containing identification information, certificate information and an officially ciphered data certificate of the certificate.

20. The method for data transfer by a standardized document structure comprising the steps of:

creating a certified resulting list by a recognition process according to the global document descriptor;

sending the certified resulting list through a communications channel;

in the receiving station the certified resulting list is checked up by an inquire to the certificate center using a secured communications channel;

on base of the certified resulting list it creates a graphical image of the document in accordance with a local descriptor of a graphical document image;

the created graphical image of the document has the same certificate power as the original document, because it can be always examined by an inquiry to the certificate center using a secured communications channel;

the data transfer can be made up transmitting a graphical document image through a communications channel and checking up this image by an inquiry to the certificate center using a secured communications channel.

21. The method for automatic processing of recognized data comprising:

automatic generation of documents from a list that is pointed from the composed nomenclature of structural document identifiers and characteristic structural identifiers according to claim 9;

using the composed nomenclature of structural document identifiers and characteristic structural identifiers according to claim 9 for determination of field links of generated documents with fields of the base document and functional relations between them;

the result of automatic processing is a certified resulting list for a document that can be transformed in a graphic image according to claim 20.

22. The method for document standardization, comprising:

writing in the document all or some graphical images and using all or some methods with goal to standardize the recognition, certifying, data transmitting or automatic proceeding of the documents in way that a document means each information combination written on whatever carrier using whatever method so that it can be read by whatever apparatus and can be interpreted as an image;

. a document in electronic form that is represented by its certified resulting list is in standard form, too.

23. Some documents of claim 22 are:

text documents;

accounting documents;

administrative and legal documents;

labels of goods;

addresses of parcels;

currency, securities, stocks, postage stamps;

printed publications - books, periodicals;

automatic or hand created forms for handmade input in a machine as game slips, cash-slip checks, voting-papers etc.;

identification documents;

graphic images in cardiograph;

electronic publications - graphic images published through electronic way;

transformed in graphic form non-graphical information processed according to one or more of claimed methods.

24. The certificate center, comprising:

means for secure keeping necessary information about creation of content certificate of documents;

means for making document certificates;

means for secure keeping information about ciphering-deciphering of a certificate and linked to it information about a registered customer or an independent certification; and means

for ciphering-deciphering; and means for extracting information about a registered customer in relation with the ciphering-deciphering information;

means for establishing secured communication channels to the registered customers, using the respective information;

means for exact determining the date and the time, when they are written in the ciphered certificate;

means for keeping and proceeding the generated information of the center about ciphering-deciphering the certificate area where is disposed the officially generated certificate of the certificate.

25. The global accessible site, that keeps the global descriptor, comprising:

means for keeping information about the global descriptor;

means for information search;

means for establishing secured communication channels to the registered customers or means for transmitting certified information according to the claim 18.

STATEMENT UNDER ARTICLE 19(1)

There are some important differences between "certified resulting list" and "resulting list" and between "certificate" and "ciphered certificate". Confusion was made using short expressions to indicate two different senses. Very important is the use of secured communications channel for secure communication between the customers and the certificate centre and between the customers and the global accessible site, that contains the global descriptor. In the last case is obvious that can be used the described certification method for information authenticity guarantee. Because of these were made many amendments in the claims. See point 9 (and all its subpoints) of "Detailed description" and point 9 (and all its subpoints) of "Detailed description of preferred embodiment" and fig.14, fig.15 for amendments of claims 1, 18, 20, 21 and for new claim 24. See point 8 of "Detailed description" and point 8 (and all its subpooints) of "Detailed description of preferred embodiment" and fig. 14 for amemndments of claim 8 and for new claim 25. Some claims were amended only to be independent; other claims were amended without change their sense, but to be clearly. These claims are 2 to 4, 10 to 15, 17, 19, 22, and 23.
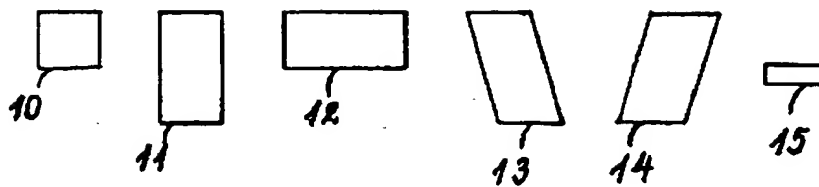
FIG.1
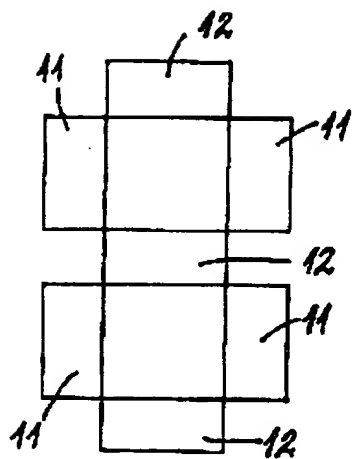
FIG.2
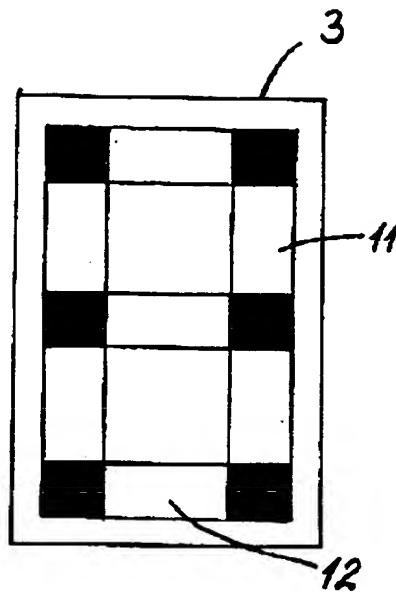
FIG.3

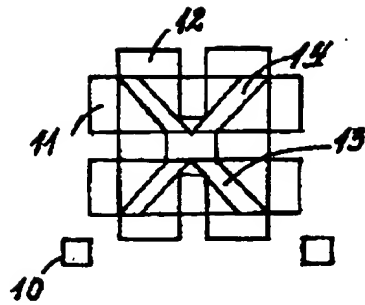FIG.4
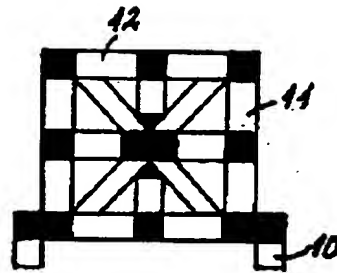
FIG.5

FIG.6
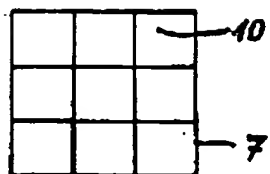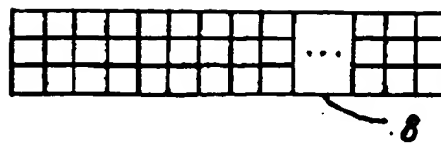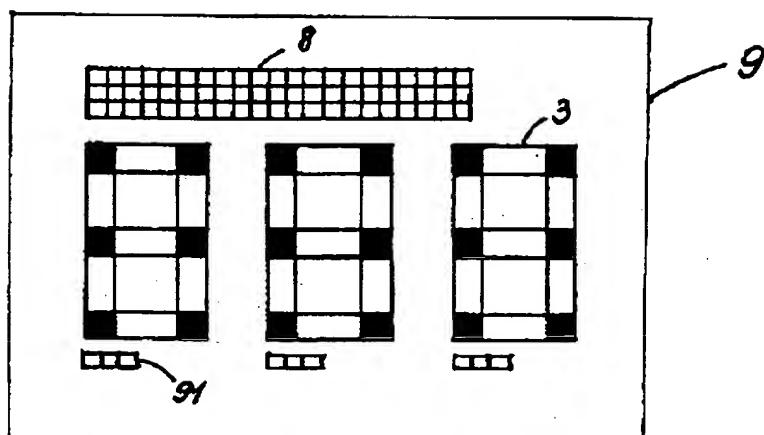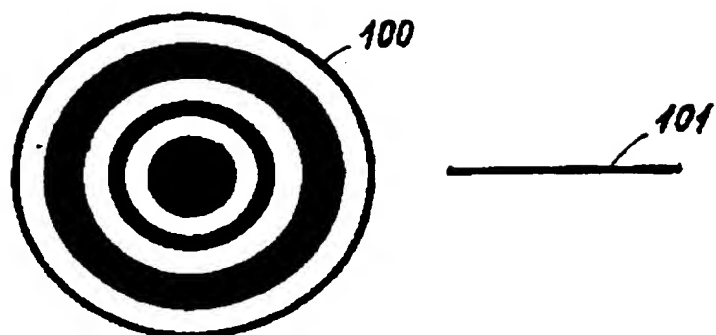
FIG.7

FIG.8

FIG.9

FIG. 10
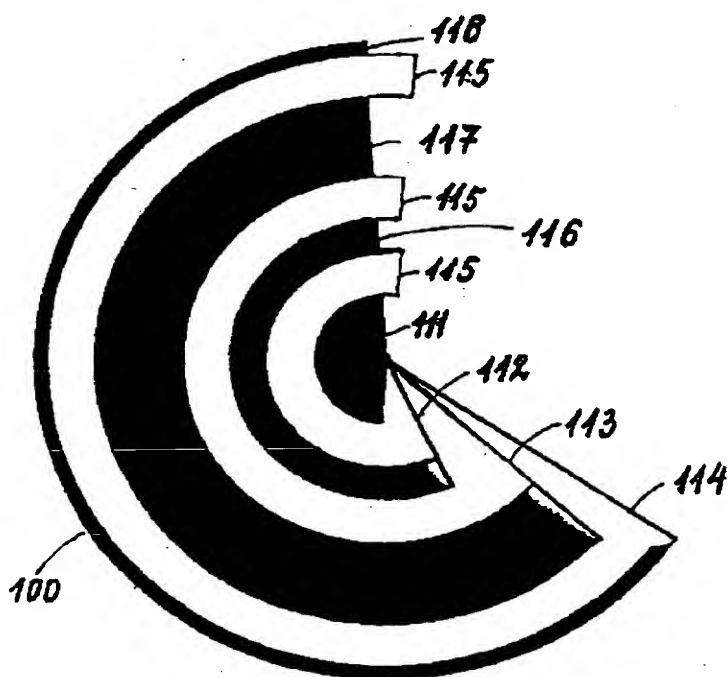

FIG. 11
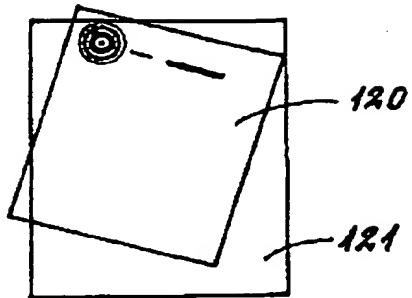
FIG. 12

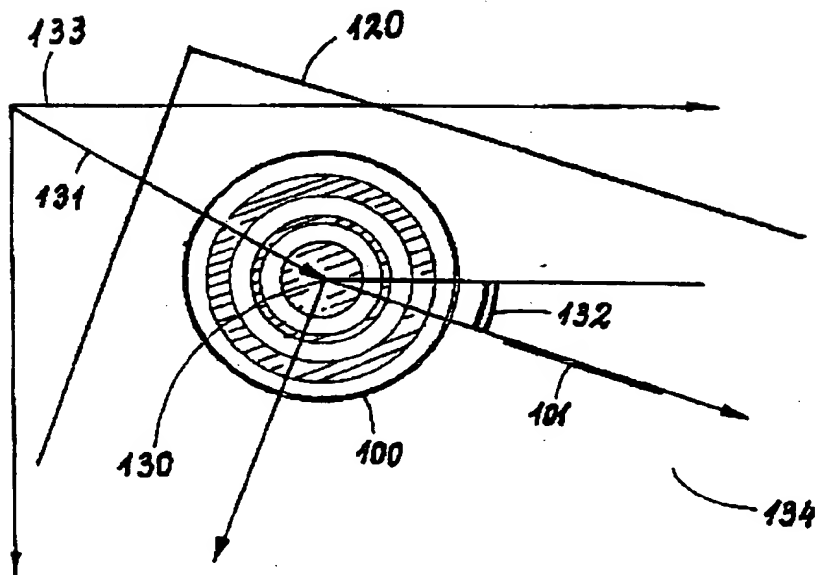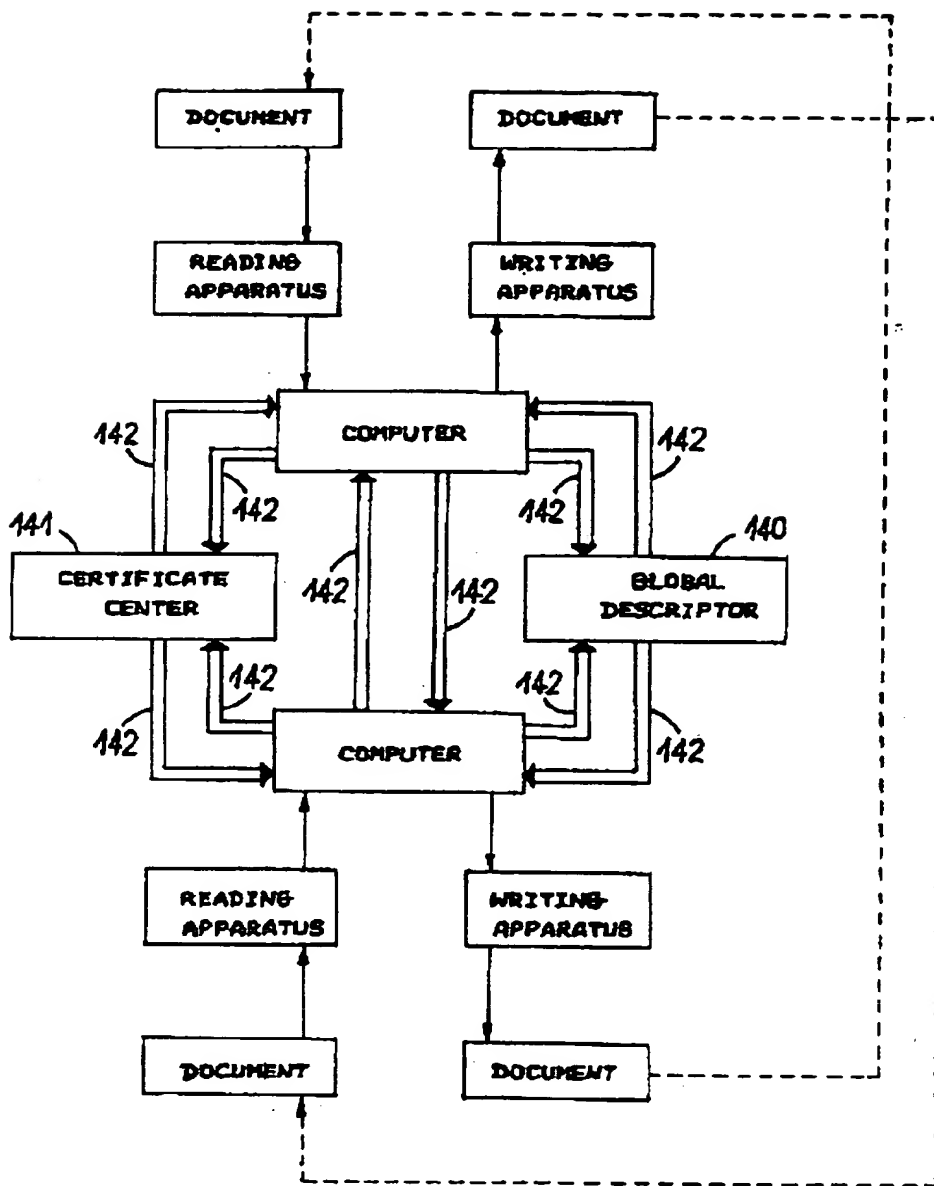FIG. 13

FIG.14

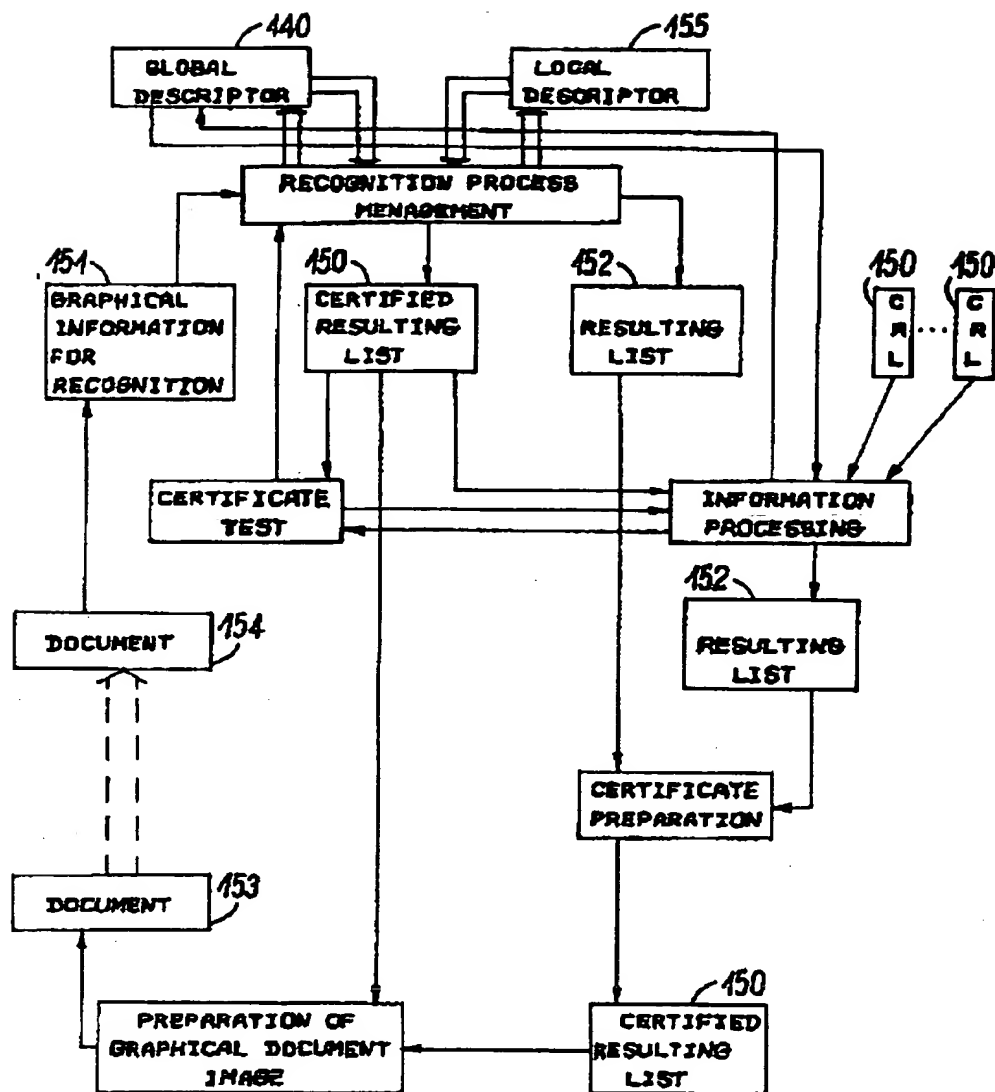*FIG. 15*

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7   G06K9/20

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7   G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

INSPEC, WPI Data, IBM-TDB, PAJ, EPO-Internal, COMPENDEX

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LOPRESTI D P ET AL: "CERTIFIABLE OPTICAL CHARACTER RECOGNITION" PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON DOCUMENT ANALYSIS AND RECOGNITION, 20 October 1993 (1993-10-20), XP000764651 the whole document	1-23
A	S. G. ADEN & AL.: "DOCUMENT Format Selection and Control Process" IBM TECHNICAL DISCLOSURE BULLETIN, vol. 26, no. 9, 1 February 1984 (1984-02-01), pages 4718-4719, XP002144726 New York, US the whole document	1-23

-/--

[X] Further documents are listed in the continuation of box C.   [X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 August 2000	24/08/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Granger, B

Form PCT/ISA/210 (second sheet) (July 1992)

1

INTERNATIONAL SEARCH REPORT

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0 386 867 A (FISCHER ADDISON M) 12 September 1990 (1990-09-12) abstract; figure 1	1-23
A	US 5 745 610 A (JOHNSON WALTER A L) 28 April 1998 (1998-04-28) cited in the application abstract; claims 22,26; figure 15	1-23
A	GB 2 102 997 A (ROUNDEL ELECTRONICS) 9 February 1983 (1983-02-09) abstract; figure 5	2

1

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
EP 0386867	A	12-09-1990	US	5005200 A	02-04-1991
			AT	113429 T	15-11-1994
			AT	150605 T	15-04-1997
			AU	620291 B	13-02-1992
			AU	4242589 A	13-09-1990
			CA	2000400 A,C	07-09-1990
			DE	69013541 D	01-12-1994
			DE	69013541 T	09-03-1995
			DE	69030268 D	24-04-1997
			DE	69030268 T	26-06-1997
			DK	386867 T	03-04-1995
			EP	0586022 A	09-03-1994
			ES	2036978 T	01-01-1995
			ES	2098651 T	01-05-1997
			GR	93300050 T	30-06-1993
			JP	2291043 A	30-11-1990
			US	5214702 A	25-05-1993
US 5745610	A	28-04-1998	EP	0635798 A	25-01-1995
			JP	7056956 A	03-03-1995
GB 2102997	A	09-02-1983	AU	8682382 A	02-02-1983
			EP	0083601 A	20-07-1983
			WO	8300243 A	20-01-1983